FİZİKSEL TIP VE REHABİLİTASYON BİLİMLERİ DERGİSİ
Journal of Physical Medicine and Rehabilitation Sciences J PMR Sci

ORIGINAL RESEARCH ORİJİNAL ARAŞTIRMA

DOI: 10.31609/jpmrs.2025-109562

How Reliable Is It to Use the Virtual Model in the Real World? Response to Stroke Questions Via ChatGPT

Sanal Modeli Gerçek Dünyada Kullanmak Ne Kadar Güvenilir? ChatGPT Aracılığıyla İnme Sorularına Yanıt

Fatma ÖZCAN[a], Merve ÖRÜCÜ ATAR[a], Yasin DEMİR[a], Eda GÜRÇAY[a]

[a]University of Health Sciences Faculty of Medicine, Ankara Gaziler Physical Medicine and Rehabilitation Training and Research Hospital, Department of Physical Medicine and Rehabilitation, Ankara, Türkiye

ABSTRACT Objective: Nowadays, artificial intelligence chatbots are increasingly used by patients to obtain information about the diagnosis, of their diseases. This study aimed to determine the reliability and usability of Chat Generative Pre-trained Transformer (ChatGPT) in stroke-related questions. **Material and Methods:** A total of 39 questions were prepared according to 3 keywords ("general information", "complications" and "rehabilitation"), which were identified as the most frequently searched keywords in Google Trends, and were evaluated simultaneously by 2 raters according on a 7-point Likert scale for reliability and usability. **Results:** Inter-rater Cronbach α scores indicated almost perfect to substantial agreement for both reliability and usability scores (α between 0.813-0.949, and α between 0.303-0.857, respectively). The highest mean reliability score was for "general information" (mean 5.2). The lowest average was for the "rehabilitation" section (mean 4.1). The "complications" for rater 1 and "general information" for rater 2 had the highest mean scores for the usability (mean 5.4), and the lowest mean value was recorded in the "rehabilitation" section (mean 4.6, 4.2, respectively). **Conclusion:** ChatGPT's responses to the stroke-related questions were reliable and useful. However, it should be kept in mind that ChatGPT may provide incorrect and incomplete information, especially in the "rehabilitation" section, which may lead to significant deficiencies in disease management.

Keywords: ChatGPT; stroke; reliability; usability; artificial intelligence

ÖZET Amaç: Günümüzde, yapay zekâ sohbet robotları hastalar tarafından hastalıkları hakkında bilgi edinmek için giderek daha fazla kullanılmaktadır. İnme ile ilgili sorularda Chat "Generative Pre-trained Transformer'ın (ChatGPT)" güvenilirliğini ve kullanılabilirliğini belirlemek amaçlanmıştır. **Gereç ve Yöntemler:** Google Trends'te en sık aranan anahtar kelimeler olarak belirlenen 3 anahtar kelimeye ("genel bilgiler", "komplikasyonlar" ve "rehabilitasyon") göre hazırlanan toplam 39 soru, 2 değerlendirici tarafından 7 puanlık Likert ölçeğine göre güvenilirlik ve kullanılabilirlik açısından eş zamanlı olarak değerlendirilmiştir. **Bulgular:** Değerlendiriciler arası Cronbach α puanları, hem güvenilirlik hem de kullanılabilirlik puanları için "mükemmel" ile önemli ölçüde uyum olduğunu göstermiştir (sırasıyla α 0,813-0,949 arasında ve α 0,303-0,857 arasında). En yüksek ortalama güvenilirlik puanı "genel bilgiler" içindi (ortalama 5,2). En düşük ortalama ise "rehabilitasyon" bölümü içindi (ortalama 4,1). Değerlendirici 1 için "komplikasyonlar" ve değerlendirici 2 için "genel bilgiler" kullanılabilirlik için en yüksek ortalama puanlara sahipti (ortalama 5,4) ve en düşük ortalama değer "rehabilitasyon" bölümünde kaydedildi (sırasıyla ortalama 4,6, 4,2). **Sonuç:** ChatGPT'nin inme ile ilgili sorulara verdiği yanıtlar güvenilir ve yararlıydı. Ancak, ChatGPT'nin özellikle "rehabilitasyon" bölümünde yanlış veya eksik bilgi sağlayabileceği ve bunun da hastalık yönetiminde önemli eksikliklere yol açabileceği unutulmamalıdır.

Anahtar Kelimeler: ChatGPT; inme; güvenilirlik; kullanılabilirlik; yapay zekâ

Correspondence: Fatma ÖZCAN
University of Health Sciences Faculty of Medicine, Ankara Gaziler Physical Medicine and Rehabilitation Training and Research Hospital, Department of Physical Medicine and Rehabilitation, Ankara, Türkiye
E-mail: fatma_yumus@hotmail.com

Peer review under responsibility of Journal of Physical Medicine and Rehabilitation Science.

Received: 14 Feb 2025 ***Received in revised form:*** 29 Apr 2025 ***Accepted:*** 07 May 2025 ***Available online:*** 15 May 2025

The development of digital technology has led to changes in many areas of healthcare. In particular, there has been an increase in recent years in the use of digital consultations, telemedicine, remote treatment and mobile health applications to improve healthcare services.[1] Artificial intelligence (AI) technology offers significant benefits in many areas, from clinical decision support systems to patient education.[2] AI was defined in the 1950s by McCarthy et al. to refer to the branch of computer science that uses machine-based approaches to make predictions that mimic what human intelligence can do in the same situation.[3] AI includes features such as natural language processing, knowledge representation, automatic reasoning, and machine learning.[4] Chatbots based on AI are being used in the field of healthcare for diagnosis, screening, monitoring of patients, and the evaluation of treatment options. In the area of health care, some chatbots are being used for different purposes, including Woebot (Woebot Health, Inc), OneRemission (Keenethics, USA), Senper (Senpertech, TR), Babylon Health (eMed Healthcare, UK), Infermedica (Infermedica, Poland), Gyant (Gyant, USA), Cancer Chatbot (Belong Life, USA)

Chat Generative Pre-trained Transformer (ChatGPT) (OpenAI, USA) is an AI chatbot with a large model of the language, developed by OpenAI, with a release date of November 30, 2022.[5] The goal of ChatGPT is to generate text that can be used for a variety of processing tasks, such as mimicking natural human language, translation, text summarization and dialog systems. It is also designed to generate responses, answer questions, write creative stories, and process sequential data in a chatbot. Therefore, ChatGPT allows the creation of coherent texts by analyzing the relationship between sentences in a text.[6]

Stroke is a chronic condition that requires regular follow-up and treatment. There has also been a 50% increase in the lifetime risk of stroke over the last 20 years and the risk is now one in 4.[7] Regular clinician follow-up is important to manage complications and ensure functional independence in this group of patients, where the incidence is so high and various complications are observed.[8] However, several barriers, such as lack of financial resources, problems accessing medical resources, lack of social support and confusion due to conflicting information from different healthcare providers, make it difficult for individuals to self-manage their chronic conditions.[9] The internet and social media platforms offer opportunities for patients and caregivers to overcome these restrictions and support patients in the self-management of their chronic conditions.[10]

To the best of our knowledge, this is the first study to determine the reliability and usability of ChatGPT for stroke. Therefore, the study had 2 aims: 1) to assess the reliability of ChatGPT for stroke-related responses and 2) to determine the usability of ChatGPT for stroke-related responses.

MATERIAL AND METHODS

Individual Google Trends (Alphabet, USA) searches were performed for stroke on August 11, 2023, and the 3 most searched keywords related to these diseases were identified. The search parameters included the whole world, the period from 2000 to the present, and the subcategory "health". In terms of results, the "most relevant" option was selected in the "related questions" section. Repeated similar keywords were excluded. The most searched keywords on Google (Alphabet, USA) have been identified by search results. "General information", "complications" and "rehabilitation" were the most frequently searched keywords related to stroke. A total of 39 questions were identified according to these keywords. These questions were prepared by all the authors under the guidance of an expert in physiotherapy and rehabilitation (EG) with more than 20 years of experience in the field of stroke. "General information" section consisted of 5 questions including definition, epidemiology, risk factors, classification and diagnosis, "complications" section consisted of 25 questions and "rehabilitation" section consisted of 9 questions.

In the ChatGPT AI chatbot's "conversation" section, the questions prepared for the stroke were written. In the ongoing chat, each question was rewritten by different users in separate sessions, so that the answer to each question in the conversation section was not affected by the previous question or answer. All answers were evaluated simultaneously by 2 independent specialists in physical medicine and rehabilitation (FÖ and MÖA), who were blinded to the

answers of the other person to avoid possible bias. ChatGPT-4 responses are from the March 14, version.

Each chatbot conversation was rated on a scale of 1-7 (which 1 is the lowest score and 7 is the highest score) for its reliability and usefulness in 2 categories.[11]

The approval of an ethics committee and informed consent were not required for this study, as related research has followed a similar pathway, as conversations were evaluated on ChatGPT and no animals/human participants were involved.[11,12]

STATİSTİCAL ANALYSİS

Statistical analysis was conducted using SPSS version 15.0 (IBM®, Chicago, USA) package program. Numerical data descriptive statistics are expressed as mean±standard deviation. Numerical variables were compared between the groups using the Mann-Whitney U test and the Kruskal-Wallis test. The strength of agreement between rater 1-2 was determined as follows: slight (≤0.20), fair (range 0.21-0.40), moderate (range 0.41-0.60), substantial (range 0.61-0.80) and almost perfect (range 0.81-1.00).[13] Inter-rater reliability was determined by calculating intraclass correlation coefficients with 95% confidence intervals in a 2-way mixed effect model. A p-value of less than 0.05 was defined as statistically significant.

RESULTS

The ChatGPT results for 39 questions prepared in line with the 3 keywords used by Google Trends were scored by 2 independent, experienced raters using a 7-point Likert scale. Table 1 and Table 2 show the reliability and usability scores and inter-rater results. The inter-rater Cronbach α scores indicated "almost perfect" to "substanstial" agreement for the reliability and usability scores (α between 0.813-0.949, and α between 0.303-0.857, respectively).

When analyzing the distribution of Likert scores in terms of question topics, it was found to be between 3-6. Based on the answers to the 39 questions, the highest reliability score (both raters; point 6) was for definition, risk factors, swallowing impairments-diagnosis, shoulder pain-general information, shoulder pain- diagnosis and mirror therapy. The highest

TABLE 1: Inter-rater reliability scores and agreement between the raters

	Rater≠1	Rater≠2	Cronbach α (95% CI lower to upper)
General Information			0.938 (0.538-0.993)
Definition	6	6	
Epidemiology	4	4	
Risk factors	6	6	
Classification	4	5	
Diagnosis	5	5	
Complications	5	5	0.823 (0.606-0.923)
Speech and language impairments			
General information	5	5	
Diagnosis	5	5	
Treatment	5	6	
Spasticity			
General information	3	4	
Diagnosis	4	4	
Treatment	6	5	
Swallowing impairments			
General information	4	4	
Diagnosis	6	6	
Treatment	5	5	
Pressure ulcers			
General information	5	5	
Diagnosis	4	5	
Treatment	4	3	
Neurogenic bladder			
General information	4	3	
Diagnosis	5	5	
Treatment	5	4	
Neurogenic bowel			
General information	5	5	
Diagnosis	6	5	
Treatment	4	5	
Shoulder pain			
General information	6	6	
Diagnosis	6	6	
Treatment	5	5	
Central post-stroke pain			
General information	6	5	
Diagnosis	4	4	
Treatment	5	5	
Rehabilitation			0.975 (0.899-0.994)
Neurophysiological exercise	4	4	
Mirror therapy	6	6	
Robot-assisted therapy	5	5	
Constraint-induced movement therapy	5	4	
EMG biofeedback	5	5	
Transcranial magnetic stimulation	3	3	
Transcranial direct current stimulation	3	3	
Virtual reality	4	4	
Orthosis	3	3	

CI: Confidence interval; EMG: Electromyography

TABLE 2: Inter-rater usability scores and agreement between the raters

	Rater≠1	Rater≠2	Cronbach α (95% CI lower to upper)
General Information			0.769(-0.584-0.968)
Definition	5	5	
Epidemiology	4	5	
Risk factors	6	6	
Classification	5	6	
Diagnosis	5	5	
Complications	5	5	0.675 (-0.141-0.825)
Speech and language impairments			
General information	6	5	
Diagnosis	6	5	
Treatment	5	5	
Spasticity			
General information	3	4	
Diagnosis	5	4	
Treatment	6	5	
Swallowing impairments			
General information	4	4	
Diagnosis	6	6	
Treatment	6	5	
Pressure ulcers			
General information	6	5	
Diagnosis	5	5	
Treatment	5	3	
Neurogenic bladder			
General information	5	3	
Diagnosis	6	5	
Treatment	6	4	
Neurogenic bowel			
General information	6	5	
Diagnosis	6	6	
Treatment	5	5	
Shoulder pain			
General information	6	6	
Diagnosis	6	6	
Treatment	6	5	
Central post-stroke pain			
General information	6	5	
Diagnosis	5	4	
Treatment	5	6	
Rehabilitation			0.923(0.301-0.975)
Neurophysiological exercise	4	4	
Mirror therapy	6	6	
Robot-assisted therapy	6	5	
Constraint-induced movement therapy	5	5	
EMG biofeedback	5	5	
Transcranial magnetic stimulation	4	3	
Transcranial direct current stimulation	4	3	
Virtual reality	4	4	
Orthosis	4	3	

CI: Confidence interval; EMG: Electromyography

usability score (both raters; point 6) was for risk factors, swallowing impairments-diagnosis, neurogenic bowel-diagnosis, shoulder pain-general information, shoulder pain-diagnosis and mirror therapy. The questions about the transcranial magnetic stimulation, transcranial direct current stimulation and orthosis from the rehabilitation section were given the lowest reliability score (3 points) by both of the raters. The lowest usability score (rater 1; point 4 and rater 2; point 3) was for the same questions.

The total scores of the 3 sections and their rating by each rater are presented in Table 3. No statistically significant difference was found between the 2 raters for both reliability and usability scores for "general information" and "rehabilitation" sections (p values for reliability; 0.841, 0.863, respectively and p values for usability; 0.421, 0.436, respectively). In the "complications" section, no statistically significant difference was found between the 2 raters in terms of reliability (p=0.801), but a statistically significant difference was found in terms of usability (p=0.005).

Looking at the mean scores of both raters separately for the three sections, the highest mean reliability score was for "general information" (mean 5,2). The lowest average was for the "rehabilitation" section (mean 4,1). The "complications" for rater 1 and "general information" for rater 2 had the highest mean scores for the usability (mean 5,4), and the lowest mean value was recorded in the "rehabilitation" section (mean 4,6, 4,2, respectively).

When comparing topics by raters, there was no statistically significant difference in either reliability or usability scores except for usability for rater 1 (p=0,231 and p=0,081 for reliability, p=0,037 and p=0,084 for usability) (Table 3).

DISCUSSION

The main finding of this study is that the ChatGPT conversational bot, with some limitations, demonstrates reliability and usefulness in the stroke-related questions. The "rehabilitation" section had the lowest mean reliability and usability scores. The questions about the transcranial magnetic stimulation, transcranial direct current stimulation and orthosis from

TABLE 3: Comparison of reliability and usefulness total ratings of the sections

	General information	Complications	Rehabilitation	p value[a]
Rater≠1				
Reliability	5.0±1.0	4.8±0.8	4.2±1.1	0.231
Usability	5.0±0.7	5.4±0.7	4.6±0.8	**0.037**
Rater≠2				
Reliability	5.2±0.8	4.8±0.8	4.1±1.0	0.081
Usability	5.4±0.5	4.8±0.8	4.2±1.1	0.084
Reliability *p*[b]	0.841	0.801	0.863	
Usability *p*[b]	0.421	**0.005**	0.436	

[a]Kruskal-Wallis test; [b]Mann-Whitney U test. Data are expressed as mean±standard deviation.

the rehabilitation section were given the lowest reliability and usability scores by both of the raters.

In this study, the highest reliability scores were for definition, risk factors, swallowing impairments-diagnosis, shoulder pain-general information, shoulder pain- diagnosis and mirror therapy. Figure 1 shows the response to ChatGPT for the risk factors of spinal injury. Analysis of the responses showed that many of the risk factors mentioned in the guidelines were mentioned, such as age, gender, ethnicity, comorbidities, smoking and alcohol.[14-16] However, as the risk factors were not divided into two groups, modifiable and non-modifiable, it was not given 7 points, but 6 points by both raters. Although it does not specify which risk factors are modifiable, it is important to emphasize the importance of modifiable risk factors and lifestyle changes at the end of the answer and to refer the patient to a healthcare professional. The highest usability scores were for risk factors, swallowing impairments-diagnosis, neurogenic bowel-diagnosis, shoulder pain-general information, shoulder pain- diagnosis and mirror therapy. Given this information, it can be assumed that the answers given by the ChatGPT to the questions about risk factors, swallowing impairments-diagnosis, shoulder pain-general information, shoulder pain-diagnosis and mirror therapy are reliable and useful.

ChatGPT's response to questions about transcranial magnetic stimulation, transcranial direct current stimulation, and orthosis were given the lowest reliability and usability scores by both evaluators. Therefore, when seeking information on these topics, patients should be aware of the answers provided by ChatGPT may contain incorrect and incomplete information.

When the means of the reliability and usability scores of the three sections were compared between the raters, it was found that there was no difference between the raters except for the usability of the "complications" section. This may be due to the 7-point Likert scale used to score the responses. When scoring with a scale that allows for more parameters and multifactorial scoring, it may be found that there is no significant difference between the raters.

In the current study, the highest mean reliability and usability scores were for "general information", while the lowest scores were for the "rehabilitation" section. This may be because rehabilitation strategies can develop and change, and that there is little information available on the internet resources because rehabilitation is less known than general information. In addition, patients should be aware of information about rehabilitation may be incomplete or missing.

Patients are known to use social media platforms such as Google (Alphabet, USA), YouTube (Google, USA), Twitter (Twitter, USA), Whatsapp (Meta, USA), Instagram (Meta, USA) for information on disease diagnosis and treatment, but AI robots such as ChatGPT are expected to increase their use in the near future. According to the current study, the finding that ChatGPT is generally reliable and useful for patients to get information about stroke supports this idea. Although there are studies in the literature on stroke with YouTube, Google, Twitter, Whatsapp, no study was found with ChatGPT. A total of 14 studies, including 5 on YouTube, 5 on Twitter, 2 on Facebook (Meta, USA), 1 on both Twitter and Facebook, and 1 on Whatsapp, were included in a systematic review that examined in-

Several risk factors contribute to the development of stroke. These factors can increase the likelihood of a stroke occurring. Here are some of the most common risk factors:

1. **High Blood Pressure (Hypertension):** Elevated blood pressure is one of the most significant risk factors for stroke. It can damage blood vessels over time, increasing the risk of a blockage or rupture.
2. **Smoking:** Smoking damages blood vessels and accelerates the buildup of fatty deposits in arteries, increasing the risk of clot formation and narrowing of blood vessels.
3. **Diabetes:** Uncontrolled diabetes can damage blood vessels and increase the risk of atherosclerosis (narrowing and hardening of arteries), which can lead to stroke.
4. **High Cholesterol:** Elevated levels of LDL cholesterol (often referred to as "bad" cholesterol) can contribute to the buildup of fatty deposits in arteries, potentially leading to blockages.
5. **Obesity:** Excess weight, especially around the waist, is associated with a higher risk of stroke, as it can lead to other risk factors such as hypertension and diabetes.
6. **Physical Inactivity:** Lack of regular physical activity is linked to various risk factors for stroke, including obesity, high blood pressure, and diabetes.
7. **Unhealthy Diet:** Diets high in saturated fats, trans fats, salt (sodium), and low in fruits, vegetables, and whole grains can contribute to the development of risk factors for stroke.
8. **Excessive Alcohol Consumption:** Heavy alcohol consumption can raise blood pressure and increase the risk of stroke. Moderation is key.
9. **Atrial Fibrillation:** This heart rhythm disorder can cause blood clots to form in the heart, which can travel to the brain and cause a stroke.
10. **Other Heart Diseases:** Conditions like coronary artery disease, heart failure, and heart valve defects can increase the risk of stroke.
11. **Family History:** Having a family history of stroke or certain cardiovascular conditions can increase your risk.
12. **Age:** The risk of stroke increases with age, particularly after the age of 55.
13. **Gender:** Stroke risk factors can vary between genders. For instance, women with certain conditions like pregnancy-related high blood pressure or oral contraceptive use may have an increased risk.
14. **Ethnicity:** Some ethnic groups may have a higher predisposition to certain risk factors, such as hypertension, which can increase stroke risk.
15. **Prior Stroke or TIA:** Having a previous stroke or transient ischemic attack (TIA) increases the risk of another stroke.

It's important to note that many of these risk factors are modifiable through lifestyle changes and medical management. Managing these risk factors can significantly reduce the likelihood of experiencing a stroke. Regular check-ups with a healthcare provider and adopting a healthy lifestyle are crucial for stroke prevention.

FIGURE 1: The answer of ChatGPT for risk factors of stroke

formation about stroke on social media.[17] Eleven studies were of "fair" quality and 3 studies were of "poor" quality. These results do not indicate that ChatGPT is more useful or more reliable than other social media tools for obtaining information about stroke. However, further studies comparing ChatGPT with other social media platforms and other AI robots are needed.

The limitations of the current study are that the sources of the answers given by ChatGPT are not known, it is not known whether the source is reliable or not, and ChatGPT may give different answers to questions written at different times. Therefore, if a biased source is used as a source, it should be kept in mind that the answers given by ChatGPT may be biased. Other limitations of the study may be the small number of raters and keywords. The advantage of the study is that the evaluators assess simultaneously, in order to avoid bias.

CONCLUSION

It was concluded that ChatGPT's responses to the stroke-related questions were reliable and useful. Although it is believed that AI robots such as ChatGPT can make it easier for patients to obtain information in cases where there is difficulty in accessing healthcare services for chronic diseases such as stroke, it should be kept in mind that ChatGPT may provide incorrect or incomplete information, which may lead to significant deficiencies in disease management. Developers of AI robots should verify that health-related data are provided from reliable sources and create regular updates on this issue.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

REFERENCES

1. Turakhia MP, Desai SA, Harrington RA. The outlook of digital health for cardiovascular medicine: challenges but also extraordinary opportunities. JAMA Cardiol. 2016;1:743-4. PMID: 27580275.
2. Gunawan J. Exploring the future of nursing: Insights from the ChatGPT model. Belitung Nurs J. 2023;9:1-5. PMID: 37469634; PMCID: PMC10353608.
3. McCarthy J, Minsky ML, Rochester N, et al. A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine. 2006;27:12-4. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904
4. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol. 2019;28:73-81. PMID: 30810430.
5. OpenAI [Internet]. Introducing CHATGPT. OpenAI © 2015-2025 [Cited: February 26, 2023]. Available from: https://openai.com/blog/chatgpt/
6. Khan RA, Jawaid M, Khan AR, et al. ChatGPT-reshaping medical education and clinical management. Pak J Med Sci. 2023;39:605-7. PMID: 36950398; PMCID: PMC10025693.
7. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. Int J Stroke. 2022;17:18-29. Erratum in: Int J Stroke. 2022;17:478. PMID: 34986727.
8. Castex A, Comte M. AVC, rééducation en soins de suite et de réadaptation [Stroke, follow-up and rehabilitation care]. Soins Gerontol. 2019;24:23-6. French. PMID: 31307686.
9. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician. 2014;60:1123-33. PMID: 25642490; PMCID: PMC4264810.
10. Arcury TA, Sandberg JC, Melius KP, et al. Older adult internet use and ehealth literacy. J Appl Gerontol. 2020;39:141-50. PMID: 30353776; PMCID: PMC6698430.
11. Uz C, Umay E. "Dr ChatGPT": Is it a reliable and useful source for common rheumatic diseases? Int J Rheum Dis. 2023;26:1343-9. PMID: 37218530.
12. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29:721-32. PMID: 36946005; PMCID: PMC10366809.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74. PMID: 843571.
14. Hoh BL, Ko NU, Amin-Hanjani Set al. 2023 guideline for the management of patients with aneurysmal subarachnoid hemorrhage: a guideline from the American Heart Association/American Stroke Association. Stroke. 2023;54:e314-e370. Erratum in: Stroke. 2023;54:e516. PMID: 37212182.
15. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2019;50:e344-e418. Erratum in: Stroke. 2019;50:e440-e441. PMID: 31662037.
16. Greenberg SM, Ziai WC, Cordonnier C, et al; American Heart Association/American Stroke Association. 2022 Guideline for the management of patients with spontaneous intracerebral hemorrhage: a guideline from the American Heart Association/American Stroke Association. Stroke. 2022;53:e282-e361. PMID: 35579034.
17. Garg D, Agarwal A, Srivastava MP, et al. Use of social media in stroke: a systematic review. Ann Indian Acad Neurol. 2023;26:206-12. PMID: 37538420; PMCID: PMC10394452.